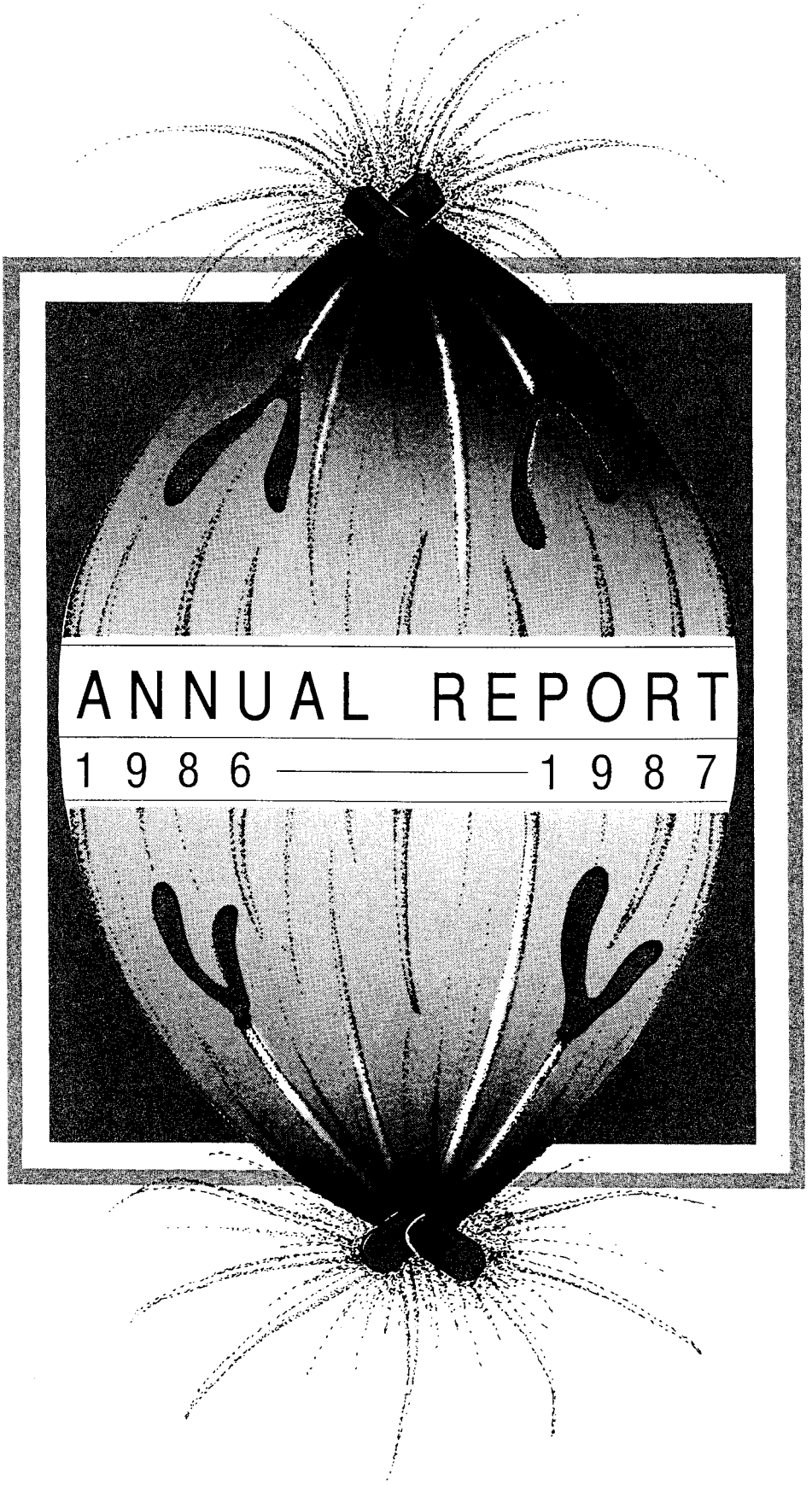
Department of

BIOCHEMISTRY AND BIOPHYSICS



ANNUAL REPORT

1986 — 1987

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

HAROLD E. VARMUS, M.D., Professor

MOLECULAR STUDIES OF RETROVIRUSES AND HEPATITIS-B VIRUSES

PAUL BATES, Postdoctoral Fellow
BRUCE BOWERMAN, Graduate Student
SHANTANU BASU, Postdoctoral Fellow
DELICIA CABALLERO, Lab Helper
MARIO CHAMORRO, Staff Research Associate
LUNG-JI CHANG, Postdoctoral Fellow
TITIA DE LANGE, Postdoctoral Fellow
HISAMARU HIRAI, Postdoctoral Fellow
JOSH KAPLAN, Graduate Student (shared with Bishop)
JAN KITAJEWSKI, Postdoctoral Fellow
TYLER JACKS, Graduate Student
HITEN MADHANI, Graduate Student
JANINE MARINOS, Administrative Assistant
JOHN MASON, Postdoctoral Fellow
EFFIE MEREDITH, Secretary
DAVID MORGAN, Postdoctoral Fellow
TUANTU NGUYEN, Staff Research Associate
SUZANNE ORTIZ, Staff Research Associate
PETER PRYCIAK, Graduate Student
MARY ANNE SCHOFIELD, Graduate Student
RAYMOND SCOTT, Lab Helper
GREG SHACKLEFORD,Postdoctoral Fellow
ANN TSUKAMOTO, Postdoctoral Fellow
JAN TUTTLEMAN, Postdoctoral Fellow
MICHAEL VERDERAME, Postdoctoral Fellow
JOHN YOUNG, Postdoctoral Fellow

In this laboratory, we use two intriguing and medically-important classes of animal viruses--the retroviruses and hepatitis B-type viruses---as points of departure for studying various aspects of the behavior of eukaryotic cells at the molecular level. Several properties of these viruses extend our concerns beyond the usual confines of virology: (i) the oncogenes of retroviruses are derived from normal cellular genes (proto-oncogenes) that are themselves targets for mutations implicated in many types of cancers; (ii) the DNA form of the retroviral genome (the provirus) is structurally similar to transposable elements from all types of living organisms; (iii) RNA-directed DNA synthesis is central to the life cycles of both classes of virus, transposition by various mobile genetic elements, and the generation of certain pseudogenes and abundantly repeated sequences in eukaryotic genomes; (iv) the complex pathological consequences of persistent infections by both virus classes involve a wide range of interactions between viral and host macromolecules; and (v) the regulation of viral gene expression displays many features characteristic of cellular genes, but more conveniently addressed with viral reagents. Consideration of such issues brings us to grips with several fundamental questions in eukaryotic biology: what is the molecular basis of cancer and of the normal regulation of growth? What is the nature of somatic mutation? What devices are used to regulate gene expression? What is the mechanism by which eukaryotic transposable elements integrate into chromosomes?

RETROVIRUSES AS TRANSPOSABLE GENETIC ELEMENTS

The Mechanism of Proviral Integration

The central event in the retrovirus life cycle is the covalent integration of a DNA copy of the RNA genome into host chromosomes. Although a great deal is known about the synthesis of unintegrated double stranded viral DNA by the virus-coded enzyme, reverse transcriptase, we know well only the structural features of integrated (proviral) DNA and few of the functional properties of the integrative mechanism. Like many transposable elements from plants, bacteria, yeast, and insects, proviruses can be found at many different sites in host genomes but are always joined to host DNA at the same sites in viral DNA. The provirus contains viral genes arranged as they are in viral RNA (most commonly: 5'-*gag-pol-env*-3'), flanked by long terminal repeats (LTRs) that are generated during reverse transcription and used for regulation of transcription. The LTRs terminate with short inverted repeats, and the entire provirus is flanked by short direct repeats of cellular origin generated during the integration step.

We are using biochemical and genetic techniques to characterize the components required for successful establishment of the provirus. A few years ago, we used site-directed mutagenesis to reveal a function (integrase) encoded by the 3' end of the *pol* gene and absolutely required for integration. More recently, we have found that synthesis and integration of retroviral DNA appear to be mediated during infection by subviral nucleoprotein complexes derived from entering virus particles; once synthesis of viral DNA is completed, these complexes can carry out the integration reaction when placed in a cell free system as described below.

Bruce Bowerman has been characterizing the nucleoprotein complexes responsible for synthesis and integration of viral DNA. He has found that cytoplasmic complexes containing linear DNA are smaller than virus particles, but sediment more rapidly (ca. 160S) than free DNA. DNA in the complexes can be precipitated with antisera against the major core (*gag*) proteins of virus nucleocapsid, and the complexes can be separated from most cellular material by density centrifugation or gel chromatography. Complexes from the nucleus of infected cells are similar in physical properties but contain closed circular DNA with one or two LTRs, as well as linear DNA.

Through a collaboration with Pat Brown, a post-doctoral fellow in Mike Bishop's lab, viral nucleoprotein complexes have been shown to mediate correct retroviral integration *in vitro*. Pat and Bruce developed a sensitive genetic assay for integration by preparing complexes from cells infected with a strain of murine leukemia virus (MLV) carrying an amber suppressor tRNA gene (supF). Extracts containing the complexes are incubated with naked lambda phage DNA bearing multiple amber mutations, and, after packaging, phage are tested for their ability to grow on a non-suppressor bacterial host.

Under these assay conditions, 0.1 to 1% of the complexes mediate correct integration of MLV DNA into bacteriophage DNA. All of the components required for the integration activity are tightly associated with the complexes, since little or no activity is lost when the complexes are prepared by gel filtration or rate zonal centrifugation. Cytoplasmic complexes are at least as effective as nuclear complexes, implying that linear DNA serves as precursor to the integrated form *in vitro*. We are uncertain whether there is a transient circular intermediate *in vitro*, but the absence of a requirement for ATP as an energy source argues against a circularization step. Development of a physical assay that allows analysis of intermediates in the reaction will now be required to examine such issues more closely.

Several other approaches are being taken to evaluate the role of the integrase and its putative viral target sequence in the integration reaction. (i) Frameshift and missense mutants of the integrase have been used as helper viruses to form viral pseudotypes with genomes carrying selectable markers. Kimiko Hagino-Yamagishi found that the rare proviruses formed by these pseudotypes are mostly aberrant and include oligomers of viral DNA generated by unknown mechanisms. (ii) The integrase coding domain has been expressed at high levels in yeast by Shantanu Basu. The protein binds non-specifically to DNA and can be separated from yeast proteins and nucleolytic activities by conventional chromatographic procedures. Preparations of integrase will be used for structural studies (in collaboration with C. Craik and R. Stroud), for attempts to complement integrase-deficient complexes in the *in vitro* integration assay, and for efforts to develop a soluble integration system. (iii) The claim that closed circular DNA is the proximal precursor to the integrated provirus depends heavily upon an experiment by Panganiban and Temin; in that experiment, the region of circular viral DNA formed by joining the ends of linear DNA was claimed to act as an integration site when relocated

at internal sites in the viral genome. We are attempting to reassess this result using the *in vitro* integration assay. (iv) Chemical inhibitors of the integrase should block integration, an essential step in the virus life cycle, and hence might be potent therapeutic agents. We are adapting the *in vitro* integration assay to screen for candidate inhibitors of the human immunodeficiency virus (HIV), the causative agent of AIDS.

Thus far we have no evidence that host factors are required for integration, but there may be a host factor that impedes integration. The mouse locus called Fv-1 is known to house dominant alleles that restrict replication of MLV at one or more early steps in the life cycle. Peter Pryciak is using susceptible and resistant cell types, in concert with the *in vitro* integration assay, to determine the mechanism by which Fv-1 alleles mediate resistance.

Retroviral Mechanisms for Gene Expression: Ribosomal Frameshifting

Retroviruses use idiosyncratic mechanisms for synthesis and integration of viral DNA, but they exploit the machinery of the host cell to perpetuate the provirus and express its genes. As described in previous reports, proviruses harbor signals for host RNA polymerase II (promoters, enhancers), for RNA processing (splicing and polyadenylation), for hormonally-regulated transcription (e.g. by glucocorticoids), and for translation. One unusual regulatory phenomenon that operates at the translational level is described in this report.

Ribosomal frameshifting appears necessary for expression of the *pol* gene of most retroviruses. Tyler Jacks first demonstrated this in efforts to explain how the Rous sarcoma virus (RSV) provirus, which contains *gag* and *pol* in adjacent, briefly overlapping reading frames, can direct synthesis of a *gag-poly* polyprotein. A spliced mRNA for this protein has not been found, suggesting that a translational mechanism might allow bypass of the *gag* terminator. Tyler used SP6 RNA polymerase to make a single RNA species from cloned RSV DNA *in vitro* and showed that the *gag-pol* RNA yields both *gag* and *gag-pol* proteins in a reticulocyte lysate, in proportions similar to those observed in infected cells. Thus it seemed likely that ribosomal frameshifting within the overlap region is used to circumvent the *gag* terminator at a fixed frequency (ca.3-5%).

Tyler then extended these observations to other retroviruses, the human immunodeficiency virus (HIV) and the mouse mammary tumor virus (MMTV). The *gag-pol* overlap in HIV is relatively long (205 nt) and the frameshift occurs near its 5' end, whereas the RSV frameshift occurs near or at the 3' end of a 58 nt overlap. MMTV (and a few other retroviruses) have an arrangement of reading frames that requires two frame shifts to enter the *pol* frame: one shift from *gag* to a reading frame for a viral protease *(pro)*, and another from *pro* to *pol*. The anticipated frameshifts occur within short overlap regions at relatively high frequency (15-25%) *in vitro* and *in vivo* to allow synthesis of adequate amounts of the full-length fusion protein. The MMTV overlap sequences are themselves unable to mediate frameshifting, implying that sequence context is important.

Tyler's recent work shows that two elements in viral RNA are essential for efficient frameshifting: a short sequence at the frameshift site and an inverted repeat (capable of forming a hairpin structure) several nucleotides downstream of the site. A few sequences that can serve as frameshift sites---U UUA, A AAA AAC, U UUA AAC, and A AAU UUU (with triplets from the initial reading frame)---were first identified by sequence gazing. Use of the U UUA site in RSV and HIV RNA has been confirmed by aminoacid sequencing of the "transframe" protein and by site-directed mutagenesis of the site. Tyler and Hiten Madhani demonstrated the importance of the potential hairpin by showing that 3' deletion mutations impair frameshifting when they invade the stem; frameshifting is also impaired by a five base mutation that destabilizes the stem and restored by a compensatory mutation. The presence of the hairpin sequence correlates with a translational pause near the frameshift site, suggesting that retardation of polyribosomes is required for slippage into the -1 reading frame. Hiten is now attempting to gauge the allowable spacing between the site and the hairpin. The tRNAs involved in the frameshift events have yet to be identified.

We have naturally sought other contexts in which frameshifting might occur. Frameshifting in the +1 direction almost certainly occurs during translation of RNA of the yeast retrotransposon Ty and perhaps during synthesis of the hepatitis B virus reverse transcriptase (see below). Robin Colgrove has written a computer program to seek genes likely to mediate -1 frameshifts based upon Tyler's consensus sequences; several candidates will soon be subjected to experimental test.

Retroviruses as Genetic Vectors

Retroviruses have demonstrated their ability to transduce cellular genes and ferry them as part of

viral genomes into new cells, where they are efficiently expressed. We and many others are exploiting this property of retroviruses to study genes (see section II), principally with MLV vectors that are suitable for expressing two cloned genes, at least one of which is selectable. These vectors can be propagated as helper-free, high titre, replication-incompetent stocks, they can be used to remove introns from cellular genes, and their proviruses can be readily recovered by molecular cloning.

The construction of cell lines expressing subsets of retroviral genes also permits close inspection of hitherto elusive viral interactions, including recombination, heterozygosis, and phenotypic mixing. Jan Tuttleman has been attempting to determine whether specific recognition signals between core and envelope components mediate the formation of virus particles; she is doing this by measuring the efficiency of formation of viral pseudotypes in which the core components are provided by RSV and the envelope proteins by MLV, RSV, or hybrid genes.

Greg Shackleford has recently succeeded in devising a vector system based upon the use of MMTV and has produced a vector that can ferry foreign genes into cells when complemented by wild-type MMTV. In addition, he has generated the first infectious clone of helper DNA, and virus produced from this clone can induce mammary tumors in mice. Components of these clones are being used by Greg and John Mason to develop an MMTV helper cell for propagation of helper-free vector stocks. In addition, mutagenesis of the helper clone will allow us to assess the hitherto elusive role of an unassigned open reading frame in the MMTV LTR during replication and tumorigenesis.

Cloning Genes for Retroviral Receptors

Entry of retroviruses into cells depends upon host-encoded transmembrane proteins that serve as virus receptors. The remarkable specificity of virus-host interactions has been known for over twenty years, from studies of the polymorphic envelope proteins of avian retroviruses and their corresponding receptors, yet little biochemical information is available about the receptors or about the nature of their interactions with viral envelope glycoproteins. Over the past few years, receptors for several animal viruses have been identified or their genes cloned; perhaps the most remarkable example is the first known receptor for a retrovirus, the lymphocyte cell surface antigen CD4 that is required for attachment of the AIDS virus, HIV, to target cells.

Paul Bates, John Young, and Suzanne Ortiz are attempting to clone the chicken genes that encode the receptors for multiple subgroups of avian retroviruses. Genomic DNA and cDNAs in expression vectors are being introduced into avian and mammalian cell lines that lack the appropriate receptors (e.g. for subgroup A or subgroup B viruses); cells that have acquired instructions for making the receptors are identified by an acquired susceptibility to avian retroviruses carrying a marker that can be selected or screened for. (These markers include antibiotic resistance genes and genes encoding cell surface markers for which good antisera are available.)

We ultimately expect these experiments to reveal the nature of the receptors, the basis for subgroup specificity and polymorphism (does it reside in protein sequence or carbohydrate modifications?), the sites of significant interaction between the receptors and the various envelope proteins, and perhaps the normal function of the host-encoded receptors.

RETROVIRUSES AS ONCOGENIC AGENTS

Retroviruses competent to induce tumors are conveniently grouped in two categories: those highly oncogenic agents that carry oncogenes transduced from normal cells and those less efficient agents that lack their own oncogenes. Among the first group are viruses employing about twenty distinctive oncogenes (v-*onc*'s); our studies have been confined almost exclusively to the most intensively examined member, v-*src*, the oncogene of Rous sarcoma virus (RSV). v-*src* encodes a phosphoprotein of 60,000 daltons (pp60v-*src*) that displays protein kinase activity *in vitro* and induces phosphorylation of tyrosine residues in several putative target proteins *in vivo*. Each v-*onc* is derived from a cellular proto-oncogene, and these in turn are often members of gene families (e.g. c-*src* is one of at least seven closely related genes, the latest of which, hck, was recently discovered by Nancy Quintrell in Mike Bishop's lab). With the exception of a few proto-oncogenes that have proved to encode growth factors or growth factor receptors, the normal functions of proto-oncogenes are unknown, though mutant (putatively oncogenic) versions of these genes frequently surface in a variety of neoplasias in man and animals.

In the second group of oncogenic retroviruses, lacking their own oncogenes, are a large number of viruses producing a wide spectrum of diseases. Over the past several years we have worked principally with the avian leukosis virus (ALV) and myeloblastosis-associated virus (MAV), which

induce B cell lymphomas and nephroblastomas, and with the mouse mammary tumor virus (MMTV), which induces mammary carcinomas. Viruses of this second type appear to act as insertional mutagens, enhancing the expression of adjacent cellular oncogenes as an initial step in tumorigenesis. The cellular protooncogenes affected by these mutations (e.g. c-myc, c-Har-ras, and int-1) have come under scrutiny as a consequence. All of our recent work in this area has been focused upon the int-1 gene activated by MMTV.

Two related issues are also briefly addressed in this section. We have begun to explore the possibility of using homologous recombination in animal cells to inactivate proto-oncogenes and thereby evaluate their normal functions. We have also initiated efforts to study genes that appear to act in a recessive manner during oncogenesis ("recessive oncogenes").

The Function of the *src* Gene of RSV: Analysis of Mutants of *src*

We have isolated a large group of non-conditional *src* mutants arising spontaneously in an RSV provirus integrated in the genome of an infected rat cell. These mutants make full sized products without protein kinase activity (presumably due to missense mutations) or truncated proteins (presumably due to nonsense or frameshift mutations), and they have proven to be useful reagents for examining the nature of eukaryotic mutations, as well as the functions of the v-*src* gene.

In the past, we have used conventional cloning and sequencing procedures to analyze selected mutations and secondary mutations that restore the wild type phenotype. Some of these have produced surprising information about a large amino terminal portion of pp60v-*src* that lies between an essential signal for myristylation at the amino terminus and the tyrosine kinase domain in the carboxyterminal half of the protein. Mary Anne Schofield has shown that large substitutions and duplications near the aminoterminus are compatible with full transforming potential, and Mike Verderame has found that loss of ability to transform rat cells (with retention of transforming potential in chicken cells) is due to loss of one codon in a highly conserved sequence upstream of the kinase domain. (The latter mutant, v-*src*-L, is discussed again below.) Earlier analysis of our mutants indicated that a second short (7 Kd) protein was expressed from v-*src* at low levels; site-directed mutagenesis by Paul Bates has shown that the second protein is not required for transformation of mammalian cells. One deletion mutant previous sequenced by Graeme Mardon suggested that the seventh amino acid encoded by v-*src* might be crucial for recognition by the mammalian myristoyl transferase; Josh Kaplan has used site-directed mutagenesis to show that lysine and arginine are functional in the seventh position, whereas asparagine is not. (The asp-7 allele is mentioned again below.)

Mike Verderame has been using the so-called "polymerase chain reaction" to amplify the v-*src* gene in over a dozen cell lines containing non-conditionally mutant genes encoding 60 Kd proteins without detectable kinase activity. Sequences of the amplified DNAs should provide a survey of lesions that inactivate the v-*src* kinase.

Genetic Approaches to Host Proteins that interact with v-*src* Protein

It is likely that several host components are essential to the function of v-*src;* among these are targets for the v-*src* kinase activity and proteins that modify the v-*src* gene product and determine its locale in the cell. We have developed three genetic assaults on the problem of isolating host genes for such factors.

a) Suppressing non-transforming *src* alleles. Josh Kaplan has developed a system for isolation of host mutants that suppress v-*src* mutations. He has built recombinant MLVs carrying the *E.coli* guanosine phosphoribosyl transferase (Ecogpt) and various *src* alleles (e.g. those that produce non-transformed or partially transformed phenotypes or that are host-dependent for transformation). After infection of HPRT- rat cells and mutagenesis, transformed cells could arise by secondary mutations in *src,* by *src*-independent host mutations, or by src-dependent host mutations; the experimental set-up permits tests that distinguish among these possibilities. In theory, DNA-mediated transformation techniques can ultimately be used to isolate any host gene in which the desired src-dependent mutations arise.

This system has been put to its first experimental test with cells infected by MLV carrying the c-*src* coding sequence. Mutagenesis of non-transformed cells producing high levels of c-*src* protein has yielded a number of transformants currently being analyzed. Additional tests (e.g. with the myristylation signal mutant of v-*src,* asp-7) are also in progress.

b) Isolating cells resistant to v-*src.* Mary Anne Schofield has used MLV-based vectors carrying v-

src and metabolic markers to build a rat cell line containing three expressed wild type v-*src* loci. She has mutated this line and selected at least one revertant to a non-transformed phenotype. In this setting, restoration of the normal phenotype is more likely to occur as a result of mutations in host genes than in viral genes. Attempts are being made to characterize the resistance phenomenon and to identify and isolate the host genes involved.

c) Suppressing host range differences. v-*src*-L appears to have lost transforming activity in rat cells, while retaining it in chicken cells. With the assumption that the protein product of this allele can distinguish between rat and chicken versions of some cellular factor(s) required for transformation, Mike Verderame has been attempting to identify the relevant chicken gene. The central strategy involves co-transfection of mammalian cells carrying the v-*src*-L allele with chicken DNA and a metabolic marker, scoring for transformed colonies.

Oncogenesis by Retroviruses Without Viral Oncogenes: MMTV and the *int*-1 Gene

Our current work on proto-oncogenes activated by proviral insertion mutations is confined to the *int*-1 gene; accounts of recently completed studies of c-*myc* and c-Ha-*ras* genes can be found in last year's report and the references.

The *int*-1 gene was discovered in 1982 when Roel Nusse, then a post-doctoral fellow in our group, used the technique known as "transposon tagging" to identify genes that serve as targets for insertion mutation during mammary tumor induction by MMTV. *Int*-1 is unrelated to known retroviral oncogenes or to any other cloned mammalian gene, and it has been highly conserved during evolution. Recently, Nusse's group in Amsterdam discovered that the Drosophila homologue of *int*-1 is the segment polarity gene, wingless. About 75% of tumors in C3H mice have MMTV insertion mutations that activate expression of *int*-1 a gene that appears to be unexpressed in all adult tissues save testis (see below). The proviruses are present on both sides of a transcriptional unit; they are generally pointed away from the transcribed region and hence act as enhancers rather than promoters of int-l.

The nucleotide sequence of int-1 cDNA predicts a protein of 370 amino acids, with a hydrophobic amino terminus containing a consensus site for signal peptidase cleavage, a cysteine-rich carboxyterminus, and four potential N-linked glycosylation sites. Tony Brown and Jackie Papkoff have examined the properties of int-1 proteins made proteins made *in vitro* or in cultured cells after introduction of a transcriptionally activated gene, using poly- and monoclonal sera raised against synthetic peptides. The primary product is subject to multiple modifications, including proteolytic cleavage and several glycosylations, producing at least five distinguishable forms of *int*-1 protein. The protein(s) appear to enter a secretory pathway according to several criteria, but we have yet to detect the protein in extracellular fluids. Further efforts are being made to understand the processing pathway, to identify the biological active product of the gene, and to seek the anticipated cell surface receptor for *int*-1 protein. To assist in these tasks, *int*-1 proteins have been produced in *E.coli* by Ann Tsukamoto and Mario Chamorro, and in insect cells infected with Baculovirus vectors by Mario.

Last year, in a collaborative study with Gail Martin's lab, we showed that expression of *int*-1 is confined to the midgestational embryo and the mature testis. Greg Shackleford has recently shown that *int*-1 RNA is present only in post-meiotic cells of the testis, most abundantly in early spermatids. Expression in the embryo is confined to the neural tube, with transcripts detectable in all portions of the developing central nervous system save that destined to become the forebrain. (*In situ* hybridizations by McMahon's group in England further restricts expression to a small subset of cells within the neural tube.) Greg is attempting to ascertain the function of *int*-1 in developing neural tissue by observing the behavior of the gene in cultured cells from neural lineages.

Last year Tony Brown developed a biological assay for *int*-1, using MLV vectors to induce a dramatic change in the phenotype of an established line of mouse mammary epithelial cells (C57MG). The infected cells form dense, refractile foci, permitting a direct quantitative assay for the gene and mutant versions of it. Ann Tsukamoto is creating a library of linker insertion mutations and John Mason is making several site directed mutations (e.g. to affect potential cleavage and glycosylation sites) to map the functional domains required for production of biologically active protein. Since *int*-1 does not render the C57MG line tumorigenic, Ann is also testing the effects of the gene in other lines and in primary mammary epithelial cells. In a collaborative effort with Rudi Grosschedl, Ann is also preparing transgenic mice carrying the *int*-1 gene governed by the MMTV LTR (as found in virus-induced mammary tumors), by a histocompatibility gene promoter (to incite expression in many tissues), and by the elastase promoter (to confine expression to the exocrine pancreas). These experiments should tell us

whether the oncogenic potential of the *int*-1 locus can be demonstrated in epithelial tissues other than breast or in non-epithelial cells; they should also provide us with excellent materials for seeking additional oncogenes that participate in mammary tumorigenesis.

Tony Brown has found that, although MLV vectors carrying *int*-1 do not transform fibroblast cell lines, infected fibroblasts induce morphological transformation of adjacent C57MG cells. This apparent "paracrine" effect could be due to secreted *int*-1 protein itself or to some other factor produced by cells under the influence of the *int*-1 gene. We are attempting to distinguish between these alternatives, since the assay is a potentially useful one in a search for the *int*-1 receptor and in purification of biologically active *int*-1 protein.

Recessive Oncogenes and Wilms' Tumor

All of the well-studied oncogenes appear to operate in a dominant fashion, yet studies of several tumor types over the past fifteen years, particularly retinoblastoma and Wilms' tumor (a nephroblastoma), indicate that inactivation of certain loci on both chromosomes can play a decisive role in tumorigenesis. Cytogenetic techniques localized some of these recessive loci to chromosomal bands (e.g. the Wilms' tumor gene [WT] to band 13 on the short arm of human chromosome 11), and a major event in the past year was the serendipitous isolation of part of the retinoblastoma (RB) locus by Dryja, Weinberg, and colleagues. However, virtually nothing is yet known about the product of the RB gene or about its normal and abnormal functions.

Titia de Lange has been taking three approaches to isolation of the WT gene: functional tests for genes that suppress the oncogenic phenotype of Wilms' cell lines, cloning of translocation breakpoints near or within 11p13, and differential cDNA cloning. The last approach is based upon the observation that some retinoblastomas cannot make RB RNA whereas others make a apparently normal transcript presumed to contain small mutations; the presence or absence of WT RNA might then be the only major difference between the RNA populations in two Wilms' tumors.

Inactivating Proto-Oncogenes: A Test Case with c-fos

Understanding the normal functions of proto-oncogenes would be easier if it were possible to make mutations that inactivate both copies of such genes in mammalian cell lines or animals. To gauge the feasibility of this sort of approach, Jan Tuttleman has begun experiments, in collaboration with Mario Capecchi's lab in Utah, that are designed to interrupt the proto-oncogene c-*fos* by homologous recombination. Clones of the first exon of the gene are being constructed to contain promoterless coding sequences that produce secreted proteins (e.g. interleukin-2) or confer a selectable phenotype; such clones will be microinjected into rat fibroblast lines shown to have two copies of c-*fos* accompanied by a second selectable marker. After further screening by Southern blotting of cells that express the promoterless sequence, we expect to identify successfully microinjected cells that have inserted the interrupted c-*fos* exon into the homologous region of the chromosome. This procedure will be performed twice to disrupt both copies of c-*fos*. To avoid lethal effects of the double knock-out, a retrovirus vector carrying an intact *fos* coding sequence will be used to infect some lineages; dependence upon the "rescuing" provirus can later be tested by selecting against another marker in the provirus, thereby curing the cell of the ancillary *fos* sequence. If successful, this method will be applied to other genes and to cells (embryonic stem cells) capable of repopulating the mouse germ line.

HEPATITIS B VIRUSES AS ANALOGUES OF RETROVIRUSES

The hepatitis B viruses of man and other animals---woodchucks, ground squirrels, and ducks--have some striking resemblances to retroviruses. First, these viruses replicate their DNA genomes through RNA intermediates, using a viral enzyme to synthesize viral DNA from an RNA template in the final phase of the life cycle. Second, the hepatitis B viruses frequently establish a chronic infection, and their genomes are sometimes integrated into the host genome, though not with the specificity for viral sequences manifest by retroviruses. Third, the human and woodchuck viruses seem to have a major role in the generation of primary hepatocellular carcinoma (hepatoma), a tumor arising after long latency and often carrying integrated viral DNA, though the viral genome is not believed to contain an oncogene.

We have been attempting to understand the intricacies of the unusual replication cycle, the peculiarities of expression of the highly compact genomes, and the mysterious connection between infection and carcinogenesis characteristic of this virus class. Much of our work on these problems over the past few years is summarized in last year's report, and additional detail can be found in the report by Don Ganem, with whom all of these

studies are jointly conducted. Discussion here is confined to relatively few topics.

Biogenesis of Hepatitis B Virus Reverse Transcriptases

Although the synthesis of hepatitis B virus DNA from an RNA template has been studied in viral cores isolated from infected livers, virtually nothing is known about the physical properties of the reverse transcriptase or about the manner in which it is generated. The enzyme is presumed to be encoded mainly by the longest open reading frame in the hepatitis B viral genome, since the predicted product of that frame shares some amino acid sequence with retroviral polymerases. However, no mRNA dedicated to translation of that frame has been identified, no effective anti-polymerase antisera are available, and the enzyme has not been obtained in soluble form from virus particles.

To generate better reagents for studying the enzyme, Lung Chang has expressed the long open reading frame of duck, rodent, and human hepatitis B viruses *in vitro* and in bacterial and baculovirus vector systems. These proteins are being used to raise effective antisera, to screen sera from infected patients and animals for titre against the reverse transcriptase, and to seek enzymatic activities.

Since all hepadnaviral genomes contain overlapping core and polymerase reading frames, it is possible that ribosomal frameshifting is used to produce core-polymerase fusion proteins during infection. Unlike the situation with retroviruses, however, the frameshift would occur in the +1 rather than in the -1 direction. We have been testing this possibility in two ways. Peter Pryciak has asked whether *in vitro* translation systems from reticulocytes or rodent livers can mediate a core-polymerase frameshift; thus far the answer appears to be negative. Lung Chang has been mutagenizing the core-polymerase overlap region of the duck hepatitis B virus genome in order to distinguish between models for production of the polymerase that involve frameshifting and those that involve translational initiation in the polymerase frame; the most decisive mutants are still under test.

Determinants of the Host Range of Hepatitis B Viruses

Hepatitis B viruses generally manifest relatively narrow host and tissue tropisms, but the viral and cellular factors responsible for these tropisms (i.e. host receptors and their viral ligands) have yet to be identified. We showed a couple of years ago that the genomes of two closely related rodent viruses, the ground squirrel and woodchuck hepatitis viruses, can be recombined *in vitro* without loss of viability. Since the woodchuck virus will not grow in ground squirrels, it should be possible to assign host range determinants to a component of the viral genome by testing several recombinants for their ability to grow in ground squirrels, an exercise started by Christoph Seeger and being extended by Joel Lavine. (Both parental viruses grow in woodchucks, so the viability of all recombinants can be checked in the permissive host.) If, as anticipated, host range is determined by pre-surface sequences, more refined mutagenesis will be performed in this region.

The Oncogenic Role of Hepatitis B Viruses

We have been attempting to take advantage of the very high incidence of hepatoma in woodchucks chronically infected with WHV, as compared with the low incidence in ground squirrels chronically infected with GSHV. In collaboration with Christoph Seeger (a former post-doctoral fellow here) and Bud Tennant at Cornell, we have found, as mentioned in the preceding section, that GSHV (and GSHV-WHV recombinants) can establish sustained, high-titre infection in woodchucks. A large cohort of newborn woodchucks has been chronically infected with GSHV to ascertain whether the high frequency of tumorigenesis is determined by virus or host. If the virus is crucial, GSHV-WHV recombinants will be used to identify the region of the viral genome that influences oncogenic potency.

We are continuing to characterize the collection of human hepatomas received some years ago from collaborators in Hong Kong. Joel Lavine is extending work begun by Teddy Fung to seek rearrangements of known proto-oncogenes and to ask whether hepatitis B viral DNA is integrated in the same region of the genome in independent tumors, as might be expected if a cellular locus was a favored target for insertional activation during tumorigenesis.

1986-1987 PUBLICATIONS

VARMUS, H.E. and BISHOP, J.M. Proteins encoded by oncogenes. *Cancer Surveys* 5: 153-158, 1986.

VARMUS, H.E. Cellular and viral oncogenes. Chapter 9 in "*Molecular Basis of Blood Diseases*" (Stamatoyannopoulos, Nienhuis, Leder, and Majerus, eds.), W.B. Saunders Co., New York, pp. 271-346, 1987.

GANEM, D. and VARMUS, H.E. The molecular biology of the hepatitis B viruses. *Ann. Rev. Biochem.* 56: 654-693, 1987.

VARMUS, H.E. Reverse transcription. *Scientific American*, in press, 1987.

SEEGER, C., GANEM, D. and VARMUS, H.E. Biochemical and genetic evidence for the hepatitis B virus replication strategy. *Science* 232: 477-484, 1986.

WESTAWAY, D., PAPKOFF, J., MOSCOVICI, C. and VARMUS, H.E. Identification of a provirally activated c-Ha-*ras* oncogene in an avian nephroblastoma via a novel procedure: cDNA cloning of a chimaeric viral-host transcript. *EMBO J.* 5: 301-309, 1986.

NOTTENBURG, C. and VARMUS, H.E. Features of the chicken c-*myc* gene that influence the structure of c-*myc* RNA in normal cells and bursal lymphomas. *Mol. Cell. Biol.* 6: 2800-2806, 1986.

JAKOBOVITS, A., SHACKLEFORD, G.M., VARMUS, H.E. and MARTIN, G.M. G.M. Two proto-oncogenes implicated in mammary carcinogenesis, *int*-1 and *int*-2, are independently regulated during mouse development. *Proc. Natl. Acad. Sci. USA* 83: 7806-7810, 1986.

PERSING, D.H., VARMUS, H.E. and GANEM, D. Antibodies to pre-S and X determinants arise during natural infection with ground squirrel hepatitis virus. *J. Virol.* 60: 177-184, 1986.

BROWN, A.M.C., WILDIN, R.S., PRENDERGAST, T.J. and VARMUS, H.E. A retrovirus vector expressing the putative mammary oncogene *int*-1 causes partial transformation of a mammary epithelial cell line. *Cell* 46: 1001-1009, 1986.

PERSING, D.H., VARMUS, H.E. and GANEM, D. Inhibition of secretion of hepatitis B surface antigen by a related presurface polypeptide. *Science* 234: 1388-1392, 1986.

SCHWAB, M., KLEMPNAUER, K-H., ALITALO, K., VARMUS, H. and BISHOP, M. Rearrangement at the 5' end of amplified c-myc in human COLO 320 cells is associated with abnormal transcription. *Mol. Cell. Biol.* 6: 2752-2755, 1986.

ENDERS, G.H., GANEM, D. and VARMUS, H.E. 5' terminal sequences influence the segregation of ground squirrel hepatitis virus RNAs into polyribosomes and viral core particles. *J. Virol.* 61: 35-41, 1987.

NOTTENBURG, C., STUBBLEFIELD, E. and VARMUS, H.E. An aberrant ALV provirus inserted downstream from the chicken c-*myc* coding sequence in a bursal lymphoma results from intrachromosomal recombination between two proviruses and deletion of cellular DNA. *J. Virol.* 61: 1828-1833, 1987.

BROWN, P.O., BOWERMAN, B., VARMUS, H.E. and BISHOP, J.M. Correct integration of retroviral DNA *in vitro*. *Cell*, 49: 347-356, 1987.

STONE, J., DE LANG, T., RAMSAY, G., JAKOBOVITS, E., BISHOP, J.M., VARMUS, H. and LEE, W. Definition of regions in human c-*myc* that are involved in transformation and nuclear localization. *Mol. Cell. Biol.* 7: 1697-1709, 1987.

SHACKLEFORD, G.M. and VARMUS, H.E. Expression of the proto-oncogene *int*-1 is restricted to postmeiotic male germ cells and the neural tube of mid-gestational embryos. *Cell*, in press, 1987.

JACKS, T., TOWNSLEY, K., VARMUS, H.E. and MAJORS, J. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. *Proc. Natl. Acad. Sci. USA* 84: 4298-4302, 1987.

QUINTRELL, N. R. LEBO, VARMUS, H., BISHOP, J.M., PETTENATI, M.J., LE BEAU, M.M., DIAZ, M.O. and ROWLEY, J.D. Identification of a human gene (HCK) that encodes a protein-tyrosine kinase and is expressed in hemopoietic cells. *Mol. Cell. Biol. 7*: 2267-2275, 1987.

SEEGER, C., GANEM, D. and VARMUS, H.E. *In vitro* recombinants of ground squirrel and woodchuck hepatitis B viral DNA produce infectious virus in squirrels. *J. Virol.*, submitted, 1987.

PAPKOFF, J., BROWN, A.M.C. and VARMUS, H.E. The *int*-1 proto-oncogene products are glycoproteins that appear to enter the secretory pathway. *Mol. Cell. Biol.*, submitted, 1987.

BROWN, A.M.C., PAPKOFF, J., FUNG, Y.K.T., SHACKLEFORD, G.M. and VARMUS, H.E. Identification of protein products encoded by the proto-oncogene *int*-1. *Mol. Cell. Biol.*, submitted, 1987.

MAJOR RESEARCH SUPPORT

Source:	National Institutes of Health
Title:	Molecular Analysis of Retroviruses and Oncogenes.
Total period of award:	August 1, 1985 to July 31, 1992.
Source:	National Institutes of Health
Title:	The Molecular Biology of Hepatitis B-type Viruses.
Total period of award:	April 1, 1985 to March 31, 1989.
Source:	National Institutes of Health
Title:	Oncogenic Potential of the Hepatitis B-type Viruses
Total period of award:	March 1, 1984 to February 28, 1988.
Source:	American Cancer Society - Research Professorship
Total period of award:	1984 - .

PERSONNEL WHO LEFT THE LABORATORY 1986-1987ÆMDNMØ

1. Anthony Brown, Assistant Professor, Department of Cell Biology and Anatomy, Cornell U. Medical College

2. Greg Enders, Medical Student, UCSF Medical School

3. Kimiko Hagino-Yamagishi, Staff Researcher, Department of Viral Oncology, Cancer Institute, Tokyo

4. Carol Nottenburg, Staff Researcher, Fred Hutchinson Cancer Research Center

5. Jacqueline Papkoff, Staff Researcher, Syntex Research

6. Christoph Seeger, Assistant Professor, School of Veterinary Medicine, Cornell University

7. Hans von Dam, Graduate Student, Department of Molecular Biology, Agricultural University in Wageningen